

Integrating and Querying Patient's Medical Data

Juha Puustjärvi¹, and Leena Puustjärvi²

¹ University of Helsinki/Department of Computer Science, Helsinki, Finland

Email: juha.puustjarvi@cs.helsinki.fi

² The Pharmacy of Kaivopuisto, Helsinki, Finland

Email: leena.puustjarvi@kolumbus.fi

Abstract— Although the original purpose of the HL7's Continuity of Care Documents (CCD) was to deliver clinical summaries between healthcare organizations, nowadays they are increasingly used for collecting patients' health documentation from various healthcare providers. Usually the collected CCD documents are organized into hierarchical structures that simplify the search of documents, e.g., grouping together the documents by episode, clinical specialty or time period. Yet each clinical document is stored as a stand-alone artifact, meaning that each document is complete and whole in itself. Considering each document only as a complete and a whole in itself also has its drawback: the efficient usage of patients' health documentation often is data centric, meaning that data should be extracted from various documents and then integrated according to specific criteria. Processing such queries requires the integration of the data of the CCD documents. In this paper, we present two ontology-based methods for the integration. Which of the methods is appropriate depends on whether the header or the whole CCD documents are based on the HL7 RIM.

Index Terms—Electronic health record, clinical documentation, HL7, CCD standard, XML, semantic web

I. INTRODUCTION

An electronic health record (EHR) describes the systematic documentation of a single patient's medical history and care across time within one particular health care provider's jurisdiction . It includes a variety of types of observations entered over time by health care professionals, recording observations and administrations of drugs and therapies, orders for the administration of drugs and therapies, test results, x-rays, and reports [1].

Patient's EHRs are often stored in several healthcare providers' EHR systems. This is a consequence of living in various places, and having many healthcare providers, including primary care physician, specialist, therapists and other medical practitioners. However, although patient's health documentation is stored in several EHR systems all health documents should be accessible for the physicians treating the patient.

The problem of patients' scattered health documentation is studied in the context of Personal Health Records (PHRs) and IHE XDS. From technology point of view these two policies differ in that whether patient's health documentation is collected together in advance or dynamically.

A personal health record (PHR) is a record of a consumer (i.e., not a record of a healthcare provider as an EHR) that includes data gathered from different sources such as from health care providers, pharmacies, insurers, the consumer, and

third parties [2, 3]. Similar to EHRs it includes information about medications, allergies, vaccinations, illnesses, laboratory and other test results, and surgeries and other procedures. In order to avoid the compatibility problems in importing data to PHRs various standardization efforts on PHRs have been done [4]. In particular, the use of the Continuity of Care Record (CCR standard) of ASTM and HL7's Continuity of Care Document (CCD standard) has been proposed for using in standardizing the structure of PHRs, although these standards were originally designed to store patient clinical summaries. From technology point of view CCR and CCD-standards represent two different XML schemas that are identical in their scope in the sense that they contain the same data elements [5].

IHE Cross-enterprise document sharing (XDS) allows health care documentation to be shared between hospitals, primary care providers, and social services [1]. Its key idea is to maintain a centralized document registry of patient's health documents. The indexing information in the registry includes metadata such as patient ID, document type, author, and the location of the actual document. The key idea is that based on the location information patients scattered health documentation can be captured dynamically. Technically the registry is based on the ebXML Registry standard. The shared documents may be DICOM images, HL7 medical summaries (CCDs), and structured laboratory reports.

PHRs and IHE XDS, as well as existing EHR systems, organize patients' health records in hierarchical structures. They also support a variety of ways for composing patient's health records, e.g., grouping together the records by episode, clinical specialty or time period. The ability to compose health records is of prime importance but it does not solve the problem of content based searching. The problem here is that the efficient usage of patients' health documentation often is data centric, meaning that data should be extracted from various documents and then integrated according to specific criteria.

For example, a physician may be interested to know the average blood pressure and/or cholesterol level during the time periods the patient was using Diovan (a drug for blood pressure). However, the computation required by such queries is not provided by the query languages that are designed to address hierarchical structures such as XML documents (e.g., XPath language and XQuery language). As a result, to find out such dependencies the physician first has to retrieve the relevant documents, and then search for the required data from the documents. Such a navigation and searching may be frustrating and time consuming.

We have studied the suitability of the Semantic Web technologies for integrating patients' health documentation. In particular, we have developed two ontology-based methods for integrating CCD documents. In the first method, the RMIM (Refined Message Information Model) diagrams [1] of the CCD documents are first translated into OWL-ontologies and then these ontologies are integrated. In the second method, the XML schemas of the CCD documents are transformed into an OWL-ontology. Which integration method is appropriate depends on whether the CCD documents are based on CDA Level 2 or CDA Level 3, i.e., whether the header or the whole CCD documents are based on the HL7 RIM.

The ontology that we have developed from the CCD documents is specified by OWL (Web Ontology Language), and is called the *CCD ontology*.

Transforming CCD documents into the format that is compliant with the CCD ontology can be done automatically. Further, the query languages developed for OWL, such as RQL and SPARQL, can be used for querying patients' health data stored in the CCD ontology.

The rest of the paper is organized as follows. First, in Section II, we give an overview of XML-based PHRs and semantic PHRs. Then, in Section III, we consider the characteristics of the clinical documents defined by the CDA standard. In Section IV, we illustrate the role of the HL7 RIM and RMIMs in specifying the structure and semantics of clinical documents. In Section V we present the structure of the CCD documents, and the three levels of the HL7 CDA documents. In Section VI, we consider the integration of CDA Level 3 documents and in Section VII the integration of CDA Level 2 document. In Section VIII, we present the architecture of the EHR-Archive, which contains the integrated CCD documents. We also give an example of transforming a CCD document in the format that is compliant with the CCD ontology. Finally, in Conclusion, we discuss the challenges of our solutions as well as our future research.

II. PERSONAL HEALTH RECORDS

A. The Use of PHRs

A personal health record (PHR) is a record of a consumer that includes data gathered from different sources such as from health care providers, pharmacies, insurers, the consumer, and third parties such as gyms [4]. It typically includes information about medications, allergies, vaccinations, illnesses, laboratory and other test results, and surgeries and other procedures.

Many studies have demonstrated that the provision of information therapy can increase compliance with treatment regimens, satisfaction with the health care provider and medical facility, and improve the ultimate health outcome for the individual. It is also turned out that patients who do not understand their treatment instructions, disease management, or prescription requirements are more likely to mishandle their health, be hospitalized more frequently, and have much higher medical costs than their more involved counterparts.

We have studied information PHRs in the context of medicinal treatment. It is turned out that most patients are not satisfied with the medical treatment information on the Web though many PHRs provide links to materials or other websites that have information about consumer's health conditions or medications [6]. In particular, they have regarded many sites to be overly commercial, or they could not determine the source of the information.

An ideal PHR would provide a complete and accurate summary of the health and medical history of a consumer [7]. It is only accessible to the consumer and to those authorized by the consumer. It is not the same as electronic health record (EHR), which is designed for use by health care providers, and which is designed to contain only patient clinical summaries [8].

B. XML-Based PHRs

A problem with XML-based PHRs is that their data is document-centric-data, i.e., they are collections of documents such as documents including lab tests, prescribed medications and illnesses. By contrast, it is turned out that PHR's effective usage often is data centric, meaning that data should be extracted from various documents and then integrated according to specific criteria. For example, a consumer may be interested to know the average blood pressure and/or blood sugar concentration (glucose level) during the time periods he or she was using Norvasc (a drug for blood pressure), or the consumer may be interested to know the cholesterol values when he or she was on a diet. Unfortunately the computation required by such queries is not provided by the query languages that are designed to address XML documents (e.g., XPath language and XQuery language).

C. Semantic PHRs

In order to allow data-centric queries on PHRs we have developed a methodology for developing and maintaining semantic PHRs. By semantic PHRs we refer to PHRs, which content is structured according to a PHR-ontology, and which are stored in a knowledge base or in a database and thus can be accessed by query languages having high expression power.

In order to simplify the PHR-ontology design process and to exploit the work done on CCR standard we have developed the PHR-ontology by transforming the XML-Schema of the CCR-standard into OWL ontology [9]. In this way we have ensured that the most relevant concepts are included in the PHR-ontology.

In capturing data into semantic PHR from different XML-based data sources requires that the data is first transformed into RDF-format that is compatible with the PHR-ontology [10]. Such transformations require that a specific stylesheet is developed for each data source. Transformed data is then delivered through the SOAP-protocol to the PHR-system, which inserts the data into appropriate PHR.

A useful feature of this kind of functionality is that it supports open healthcare systems, as the exchange RDF-

data does not require the hardcoding of the communicating systems as the messages itself includes their semantic. In contrast with HL CDA-compliant XML-based messaging represents closed systems as the semantics of the messages is hardcoded to the communicating systems, i.e., only CDA-compliant XML-messages can be exchanged, and the introduction of the new message type requires the coding of the communicating systems.

D. The Platforms of PHRs

A way for classifying current PHRs is to consider the platform by which they are delivered. In paper-based PHRs health information is recorded and stored in paper format, and so the information is accessible without the need for a computer or any other devices. On the other hand, paper-based PHRs may be difficult to update and share with others. They are also subject to physical loss and damage.

In portable-storage PHRs health information is stored on a portable-storage device such as CDROM or USB flash drive. Similar to paper-based PHRs they are subject to physical loss. However their main disadvantage is that reading and updating them by the computers in healthcare organizations such as in hospitals and physician offices has turned out to be problematic.

In PC-based PHRs health information is recorded and stored in personal computer-based software that may have the capability to import data from other sources such as a hospital laboratory or physician office. PC-based health information can be copied and shared with anyone who has a compatible word processor.

In Internet-based PHRs health information is stored at a remote server, and so the information can be shared with health care providers. Many Internet-based PHRs also provide links to materials or other websites that have information about consumer's health conditions or medications. Some PHRs also provide added-value services such as multidrug interaction checking or electronic messaging between patients and healthcare providers. Internet-based PHRs also have the capacity to import data from other information sources such as a hospital laboratory and physician office. However, importing data to PHRs from other sources requires standardization. If the format of the data in PHRs and in other sources such as in EHRs do not coincide (i.e., is not standardized) then the physicians and their office staff may have to type and re-enter data into PHRs.

III. CLINICAL DOCUMENTS

Clinical documentation is used to describe the care provided to a patient. The supposition is that "if it is not documented, then it did not happen".

Clinical documents have two main functions: First, they communicate relevant clinical information between healthcare providers. Second, they support compliance with regulations set by local healthcare authorities. In addition, clinical document must be credible meaning that it is produced by a trusted authority and is itself a trusted document of care that

was provided.

The properties of clinical documents are noteworthy as they restrict the ways the documents can be stored in files and databases. Thus, also our developed solutions for capturing data from clinical documents for integration may not contradict the properties clinical documents.

For example, storing clinical documents in databases is challenging as databases are not originally designed for storing documents but rather for rapid search of data which are updated by various people. Besides, the person who uses the data does not know who entered them, and in the absence of contextual information makes it difficult to evaluate whether or not the retrieved data can be relied on.

In contrast, each clinical document must be stored as a stand-alone artifact meaning that it must include metadata that states who created it, for whom, when and where, and about what subject [1]. The author of the document determines its content and is responsible for the content. Thus, if there is any doubt about interpreting the content, the author of the document can be contacted.

The functions and requirements of clinical documents led to the six characteristics defined by the CDA standard. These are persistence, stewardship, authentication, context, wholeness and human readability.

Persistency means that a clinical document exists in an unaltered state for a time period defined by local and regulatory requirements. Hence every document has a life cycle: it is created, used and in the end destroyed.

Stewardship means that the name of the steward organization is recorded as of the time the document is created. Naturally there may be organizational changes during the life cycle of the document. However, it is not required that the history of organizational changes is recorded and maintained. Instead the knowledge of the original steward organization is sufficient to locate any subsequent organization that would retain the original copy of the document.

Authentication means that each document may be signed, physically or electronically. Clinical documents are usually signed by a clinician who takes responsibility for the content of the document. However, as more health information systems are developed that automatically produce clinical documents, this requirement is often relaxed for automatically produced documents.

Context complete the clinical document by the background associated with the document. Context information is stored in the header of the document. It makes it easier for others to use the document outside the immediate purpose for which it was created. Context information is also used in grouping documents in hierarchical structures and retrieving documents. For example, document identifier, dates and times associated with the document, the type of the document, the legal authenticator, patient whose care is described are typical items of the context.

Wholeness is a principle which does not require the whole content of the document to present to make use of single statement in the document. Instead, it requires that the single statements that are stored in different systems or files should

contain a reference back to clinical document from which they came. This is also the way how we enforce this characteristic in our used document integration strategies.

Human readability of clinical documents is required as the documents are intended to communicate information between healthcare providers who are humans. It means that even when there is coded machine-readable information within the same statement there must also be a way to display the content of the document in a way that will allow a human to read it.

IV. CONTINUITY OF CARE DOCUMENT AND HL7 CDA

A. HL7 CDA

The HL7 Clinical Document Architecture (CDA) is standard XML format for clinical documents. It is based upon HL7 Version 3 Reference Information Model (RIM). It is the UML model for healthcare information. In particular, HL7 RIM specifies the grammar of V3 messages and, specially, the basic building blocks of the language, their permitted relationships, and allowed data types.

The RIM is based on two key ideas [1]. The first idea is based on the consideration that most healthcare documentation is concerned with “happenings” and things (human or other) that participate in these happenings in various ways. Happenings have a natural life cycle such as the concept itself, an intent for it to happen, the happening, and the consequences of its happening.

The second idea is the observation that the same people or things can perform different roles when participating in different types of happening, e.g., a person may be a care provider such a physician or the subject of care such as patient.

As a result of these ideas the RIM is based on a simple backbone structure, involving three main classes, Act, Role, and Entity, linked together using three association classes ActRelationship, Participation, and RoleLink (Fig. 1).

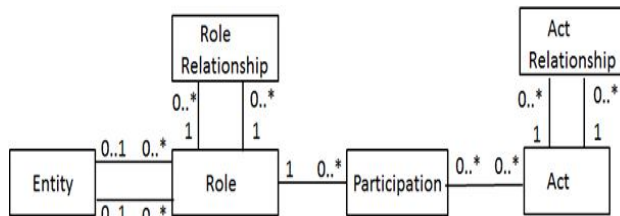


Figure1. RIM backbone structure

Each happening is an Act and it may have any number of Participations, which are Roles, played by Entities. An ACT may also be related to other Acts via Act-Relationships. Act, Role and Entity classes have a number of specializations (subclasses), e.g., Entity has a specialization LivingSubject, which itself has a specialization Person.

The classes in the RIM have structured attributes which specify what each RIM class means when used in a message (document). The idea behind structured attributes is to reduce the original RIM from over 100 classes to a simple backbone of six main classes.

B. Modeling Messages by Constrained Information Models

The RIM is not a model of healthcare, nor is it a model of any message, although it is used in messages. The structures of messages are defined by constrained information models. The most commonly used constrained information model is the Refined Message Information Model (RMIM) [1]. Each RMIM is a diagram that specifies the structure of an exchanged message.

A RMIM diagram is specified for a specific use case. The diagram is derived from the RIM by limiting its optionality. Such specifications are called CDA Profiles.

In developing a profile the RIM is constrained by omission and cloning. *Omission* means that the RIM classes or attributes can be left out. Note that all classes and attributed that are not structural attributes in the RIM are optional, and so the designer can take only the needed classes and attributes. *Cloning* means that the same RIM class can be used many times in different ways in a profile. For example, Patient and Employee are specializations of Role, and so they may both appear in the same profile.

The *multiplicities* of associations and attributes in the profile are constrained in terms of *repeatability* and *optionality*. Further, code binding is used for specifying the allowable values of the used attributes. In these constraints are specified in an XML schema. That is, the structure of CDA message such as CCD document (XML document) is specified by its XML schema and its semantics is specified by its profile, which is derived from the RIM.

A problem here is that, although the semantics of all CDA documents is tractable through a RMIM back to the RIM, we neither can use the RMIM nor the RIM in formulating queries on patient's health documentation as there are no query languages specified for the information model used in the RMIM and RIM schemas.

To illustrate the relationships of the RIM and RMIM consider the RMIM diagram of Fig. 2.

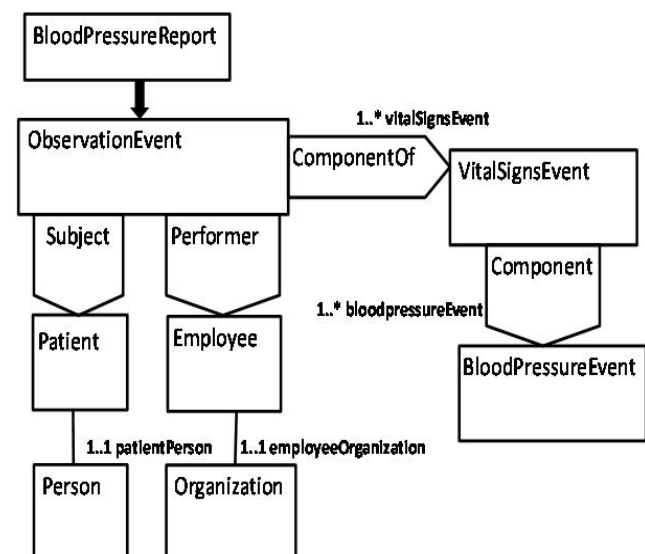


Figure 2. The RMIM of the blood pressure report.

At first note that HL7 uses its own representation of UML in RMIM diagrams: each class has its own color and shape to represent the stereotypes of these classes, and they only connect in certain ways.

The diagram specifies a blood pressure report. Its body includes the Vital signs section of the CCD. The use case behind this RMIM diagram is to exchange and store patient's blood pressures (SystolicBloodPressure and DiastolicBloodPressure) and the time of the measurement (EffectiveTime). These are the attributes of the clone BloodPressureEvent but we have omitted them, as well as other attributes, from the RMIM diagram.

The entry point of this diagram (BloodPressureReport) is ObservationEvent, which is specialization of the RIM class Act. Also classes VitalSignsEvent and BloodpressureEvent are specializations of the class Act. Classes Patient and Employee are specializations (subclasses) of the RIM class Role. Person and Organization are specializations of the RIM class Entity. Subject and Performer are specializations of the association class Participation. Component and ComponentOf are specializations of the association class ActRelationship.

V. CONTINUITY OF CARE DOCUMENTS

A. The Structure of the Continuity of Care Documents

The Continuity of Care Document (CCD) is an electronic document exchange standard for sharing patient summary information. Such summaries include the most commonly needed pertinent information about current and past health status in a form that can be shared by all computer applications, including web browsers, electronic medical record (EMR) and electronic health record (EHR) software systems.

The CCD specification is a constraint on the HL7 CDA standard. The CCD standard has been endorsed by HIMMS (Healthcare Information and Management Systems Society Though) and HITSP (Healthcare Information Technology Standards Panel) as the recommend standard for exchange of electronic exchange of components of health information.

Although some suggest that the CCD standard competes with the Continuity of Care Record standard, HL7 considers the CCD standard an implementation of the CCR standard.

Although the original purpose of the CCD documents was to deliver clinical summaries between healthcare organizations, nowadays it increasingly used for other types of messages: it is increasingly considered as set of templates because all its parts are optional, and it is practical to mix and match the sections that are needed.

Each CCD document have one primary purpose (which is the reason for the generation of the document), such as patient admission, transfer, or inpatient discharge [2]. Each CCD document, as well all CDA documents, is comprised of the Header and the Body, The sections that can appear in the Head and in the Body in a CCD document are presented in Fig. 3.

A simplified CCD document including a header and the Medications section from the Body is presented below.

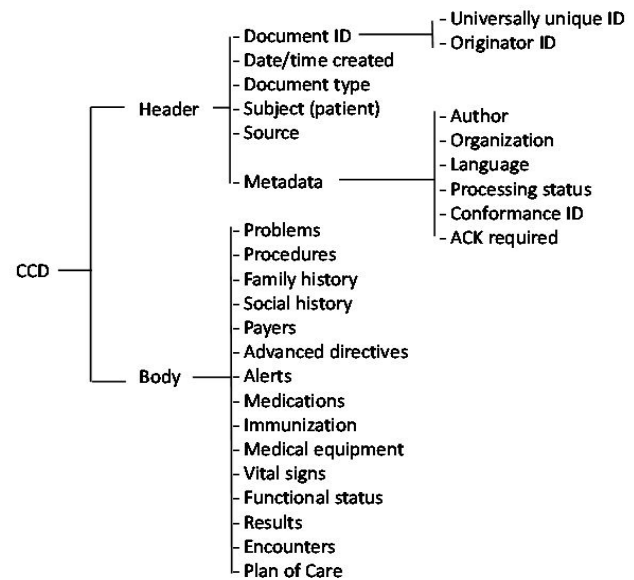


Figure 3. The structure of the CCD.

```
<SimplifiedCCDfile>
  <DocumentID>DOC_123</DocumentID>
  <Patient>
    <PatientID>AB-12345</PatientID>
    <PatientName>Tim Jones</PatientName>
  </Patient>
  <Medications>
    <Medication>
      <MedicationID>Medication.567</MedicationID>
      <DateTime>
        <ExactDateTime>2012-03-01T01:00:00</ExactDateTime>
      </DateTime>
      <Source>
        <Actor>
          <ActorID>Pharmacy of Kaivopuisto</ActorID>
          <ActorRole>Pharmacy</ActorRole>
        </Actor>
      </Source>
      <Description>
        <Text>One tablet three times a day</Text>
      </Description>
      <Product>
        <ProductName>Voltaren</ProductName>
        <BrandName>Diclofenac</BrandName>
      </Product>
      <Strenght>
        <Value>50</Value>
        <Unit>milligram</Unit>
      </Strenght>
      <Quantity>
        <Value>30</Value>
        <Unit>Tabs</Unit>
      </Quantity>
    </Medication>
  </Medications>
</SimplifiedCCDfile>
```

B. CDA Levels

The CCD document as well as any CDA documents is based on Level 1, Level 2 or Level 3. These levels differ in that whether their header and body components are based on the RIM, i.e., have a RMIM derived from the RIM:

- *Level 1:* Includes the CDA Header, which is based on a RMIM and a body consisting of an unstructured

- blob, such as PDF, DOC, or even a scanned image.
- *Level 2*: Includes the CDA Header, which is based on a RMIM, and an XML-coded body that is not based on a RMIM
- *Level 3*: Includes RMIM-based Header and Body. Thus, Level 3 documents can be automatically processed by machines.

One of the useful features of the CDA levels is that healthcare organizations can start simply by Level 1 or Level 2, and then evolve over time. The lower levels of CDA provide rather low technical requirements to adoption, while providing a migration route towards structured and semantic documents.

VI. INTEGRATING CDA LEVEL 3 CCD DOCUMENTS

A. Transforming RMIM Diagrams into OWL

Although the semantics of all CDA documents is tractable through a RMIM back to the RIM, we neither can use the RMIM nor the RIM in formulating queries. The reason is twofold: First, each RMIM diagram only models one type of document. Second, there are no query languages specified for the information model used in the RMIM and RIM schemas.

For these reasons we first transform RMIM diagrams into OWL, and then integrate these OWL-ontologies. The result of the integration is the CCD ontology. As it is OWL ontology we can define data centric queries by the query languages, such as RQL and SPARQL, which are developed for OWL ontologies.

Transforming a RMIM diagram into OWL is straightforward in the sense that the both models are object-oriented although the notation used in RMIM diagrams slightly differs from the traditional UML notation. Yet their basic modeling primitives are the same, namely classes, subclasses, properties and values. The classes are also connected in a similar way through properties.

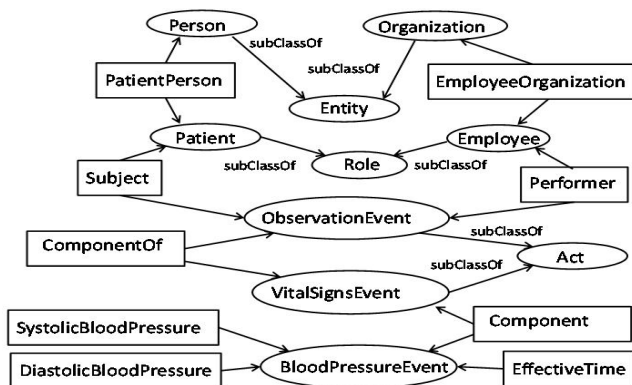


Figure 4. Graphical presentation of the RMIM ontology BloodPressureReport.

The RMIM diagram of Fig. 2 is presented in OWL in a graphical way in Fig. 4. In this graphical representation ellipses represent classes and subclasses while rectangles represent data type and object properties. Classes, subclasses, data properties and object properties are modeling primitives in

OWL. Object properties relate objects to other objects while datatype properties relate objects to datatype values. Note that, in Fig. 4 we have attached datatype properties only to the class BloodPressureEvent.

A portion of the graphical RMIM ontology of Fig. 4 is presented in OWL as follows:

```
<rdf:RDF
  xmlns:rdf=http://www.w3.org/1999/02/22-rdf-syntax-ns#
  xmlns:rdfs=http://www.w3.org/2000/01/rdf-schema#
  xmlns:owl=http://www.w3.org/2002/07/owl#>
  <owl:Ontology rdf:about="ccdOntology"/>
  <owl:Class rdf:ID="act"/>
  <owl:Class rdf:ID="role"/>
  <owl:Class rdf:ID="entity"/>
  <owl:Class rdf:ID="observationEvent">
    <rdfs:subClassOf rdfs:resource "#act"/>
  </owl: class>
  <owl:Class rdf:ID="vitalSignsEvent">
    <rdfs:subClassOf rdfs:resource "#vitalSignsEvent"/>
  </owl: class>
  <owl:Class rdf:ID="bloodPressureEvent">
    <rdfs:subClassOf rdfs:resource "#act"/>
  </owl: class>
  <owl:ObjectProperty rdf:ID="component">
    <rdfs:domain rdf:resource="#vitalSignsEvent"/>
    <rdfs:range rdf:resource="#bloodPressureEvent"/>
  </owl:ObjectProperty>
  <owl:DatatypeProperty rdf:ID="systolicBloodPressure">
    <rdfs:domain rdf:resource="#bloodPressureEvent"/>
    <rdfs:range rdf:resource="&xsd:integer"/>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="diastolicBloodPressure">
    <rdfs:domain rdf:resource="#bloodPressureEvent"/>
    <rdfs:range rdf:resource="&xsd:integer"/>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="effectiveTime">
    <rdfs:domain rdf:resource="#bloodPressureEvent"/>
    <rdfs:range rdf:resource="&xsd:time"/>
  </owl:DatatypeProperty>
  .
  .
  .
</rdf:RDF>
```

B. Integrating RMIM Ontologies

In the development of the CCD ontology we have first translated RMIM ontology into OWL. Then this ontology (the CCD ontology) is extended step by step by integrating other RMIM ontologies with the ontology. Hence the CCD ontology is incremental: when a new CCD document type (RMIM) is introduced, the CCD-ontology is extended accordingly.

Each integration step is comprised of two successive phases: In the first phase the CCD ontology is merged with the CCD ontology, and then in the second phase potential conflicts are detected and resolved.

To illustrate the merging phase, consider the CCD document (named MedicationReport), which RMIM diagram is presented in Fig. 5, and the graphical OWL ontology derived from this RMIM diagram is presented in Fig. 6.

In merging phase, we add those elements (classes, object properties and datatype properties) from the Medication report ontology to the Blood pressure report ontology that do not include in both ontologies. Such classes are SubstanceAdministration, Manufactured Product, and LabeledDrug. Correspondingly such object properties are

Consumable, Manufactured Product and Manufactured Organization.

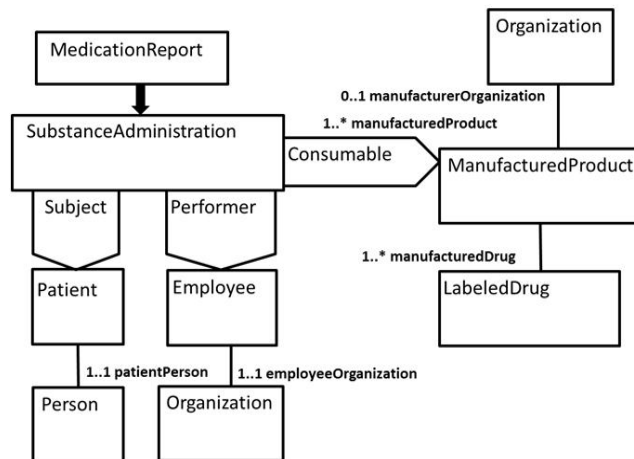


Figure 5: The RMIM of the medication report.

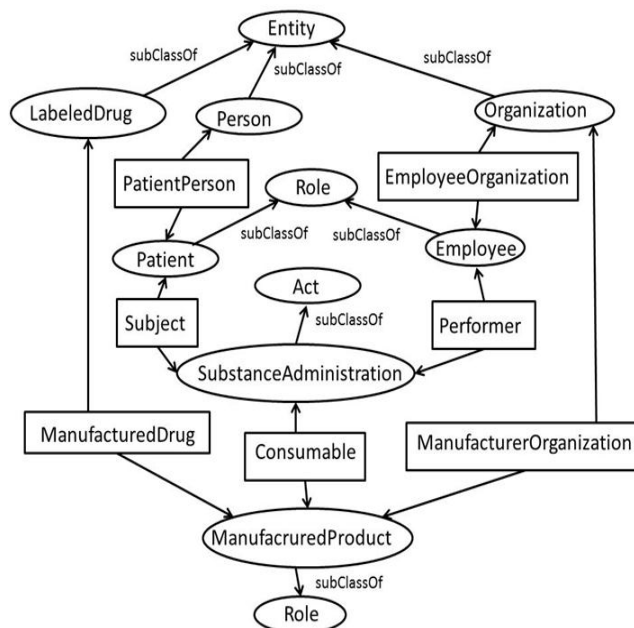


Figure 6: Graphical presentation of the RMIM ontology MedicationReport.

In merging phase, we add those elements (classes, object properties and datatype properties) from the Medication report ontology to the Blood pressure report ontology that do not include in both ontologies. Such classes are SubstanceAdministration, Manufactured Product, and LabeledDrug. Correspondingly such object properties are Consumable, ManufacturedProduct and Manufactured Organization.

Note that in graphical OWL representations (for simplicity) we have specified only a few datatype properties, and so our used examples do not reveal the datatype properties that we should insert in the integrated ontology (CCD ontology).

However, assuming that clone (class) Person has the datatype property JobTitleName in the Medication report ontology but not in the Blood Pressure report ontology, then the datatype property JobTitleName should be inserted into

the integrated ontology. So, in the merging phase we have to insert the above OWL code to the following ontology code.

```
<owl:Class rdf:ID="labeledDrug">
  <rdfs:subClassOf rdf::resource "#entity"/>
</owl: class>
<owl:Class rdf:ID="substanceAdministration">
  <rdfs:subClassOf rdf::resource "#act"/>
</owl: class>
<owl:Class rdf:ID="manufacturedProduct">
  <rdfs:subClassOf rdf::resource "#role"/>
</owl: class>
<owl:ObjectProperty rdf:ID="manufacturedDrug">
  <rdfs:domain rdf:resource="#manufacturedProduct"/>
  <rdfs:range rdf:resource="#labeledDrug"/>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="manufactureeOrganization">
  <rdfs:domain rdf:resource="#manufacturedProduct"/>
  <rdfs:range rdf:resource="#organization"/>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="consumable">
  <rdfs:domain rdf:resource="#substanceAdministration"/>
  <rdfs:range rdf:resource="#lmanufacturedProduct"/>
</owl:ObjectProperty>
<owl:DatatypeProperty rdf:ID="jobTitleName">
  <rdfs:domain rdf:resource="#person"/>
  <rdfs:range rdf:resource="xsd:string"/>
</owl:DatatypeProperty>
```

After this we have to detect and resolve conflicts. However, in the context of RMIM ontologies detecting and resolving conflicts is not as complex as in general: the “backbone structure” of the RIM ensures that the same concept has the same semantics in all RMIM ontologies. The only sources of heterogeneity arise from constraining the classes (clones) in different ways. We next consider these cases and specify how we have resolved these conflicts.

As a result of omission a class (clone) appearing in two RMIMs may be constrained in the way that their attributes are not the same, and so the corresponding RMIM ontologies are heterogeneous. We have solved such heterogeneity by taking the union of attributes, i.e., if an attribute appears in a clone of a RMIM then the attribute also appears in the integrated ontology.

Most associations and attributes in the RIM allow repeatability. That is, attributes and associations can be constrained by making such multiplicities mandatory, non-repeatable optional, or multivalued.

We have solved the heterogeneity caused by multiplicity by constraining the integrated ontology by the constraint, which is the disjunction (“union”) of the repeatability constraints originated from the integrated RMIM ontologies.

A data type conflict arises if the same attribute of the same class in two different RMIM diagrams is typed in different ways. However, such conflicts are outside the scope of OWL as it can only use the datatypes allowed in OWL. On the other hand, each RMIM diagram has a corresponding XML schema specifying the structure of the XML documents that are compliant with the RMIM diagram, and the data types of its attributes and elements are specified by the data types supported by the XML Schema (XML language). In the CCD

ontology we do not have to change these specifications as XML data types can be used in OWL.

With code binding we have assumed that though the same attribute or simple element in two clones does not use the same code binding (i.e. use different coding systems, e.g., Fahrenheit and Celsius) their XML specific data types are still equal, e.g., decimal.

VII. INTEGRATING CDA LEVEL 2 CCD DOCUMENTS

A. Extracring Elements from XML Schemas

The first step in ontology development process is to find the relevant terms that should appear in the ontology. In this stage we have exploited CCD files. In transforming the XML schema of the CCD file to OWL-ontology [14] we have used the following rules [10]:

1. The complex elements of the XML-schema are transformed into OWL classes.
2. The simple elements of the XML-schema are transformed into OWL data properties such that the complex element is the domain of the data properties.
3. The attribute of the XML-schema are transformed into OWL data properties.
4. The relationships between complex elements must be named and transformed to OWL object properties.

In order to illustrate these rules consider the graphical OWL-ontology in Fig. 7, which is derived from the elements presented in the code of the CDD document. It is a simplified CCD document including a header and the Medications section.

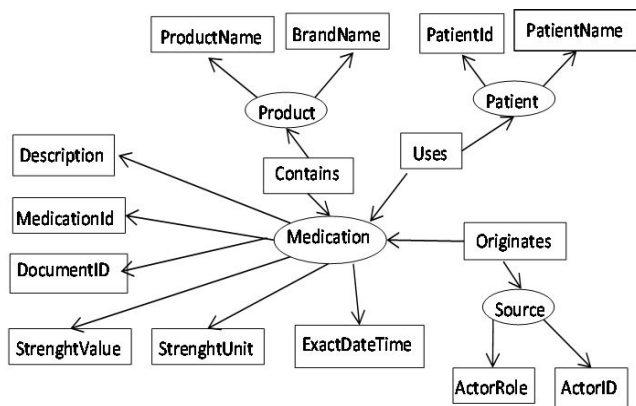


Figure 7. A simple graphical EHR-Ontology.

Note that as the OWL does not support structured attributes we have not transformed all complex elements to classes but rather the complex elements that do not have identification have been transformed to a set of properties. For example the following complex element Strenght of Fig. 2

```
<Strenght>
  <Value>50</Value>
  <Unit>milligram</Unit>
</Strenght>
```

is first transformed into data properties StrenghtValue and StrenghtUnit, and then connected to the OWL class Medication.

Note also that as a result of the transformation rule 4, we have inserted the object properties Originates, Uses, and Contains in the profile ontology. As a matter of fact whether we have to apply rule 4 depends on whether the CCD documents are based on CDA Level 2 or CDA Level 3. In the former the body of the document is comprised of XML coded sections that can be rendered in human readable form, while in the latter the sections are encoded using the HL7 V3 Clinical Statement pattern [1], i.e., are based on the RIM.

The point here is that in the RIM the classes are linked together using association classes (see Fig. 1), which have the same semantics as object properties in OWL. So the difference between CDA Level 2 and CDA Level 3 is that the former does not capture semantics while the latter does and so no extra semantics (object properties) are not needed to insert in the OWL-ontology.

In order to illustrate OWL-ontologies, the graphical ontology of Fig. 10 is presented in OWL in Fig. 11. Due to the space limits, we have omitted the specifications of the data properties such as PatientName, ProductName and BrandName.

```

rdf:RDF
  xmlns:rdf=http://www.w3.org/1999/02/22-rdf-syntax-ns#
  xmlns:rdfs=http://www.w3.org/2000/01/rdf-schema#
  xmlns:owl=http://www.w3.org/2002/07/owl#>
  <owl:Ontology rdf:about="ProfileCCDontology"/>
  <owl:Class rdf:ID="Patient"/>
  <owl:Class rdf:ID="Medication"/>
  <owl:Class rdf:ID="Source"/>
  <owl:Class rdf:ID="Product"/>
  <owl:Class rdf:ID="LabTest"/>
  <owl:ObjectProperty rdf:ID="Uses">
    <rdfs:domain rdf:resource="#Patient"/>
    <rdfs:range rdf:resource="#aMedication"/>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:ID="Contains">
    <rdfs:domain rdf:resource="#Medication"/>
    <rdfs:range rdf:resource="#Product"/>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:ID="Originates">
    <rdfs:domain rdf:resource="#Medication"/>
    <rdfs:range rdf:resource="#Source"/>
  </owl:ObjectProperty>
</rdf:RDF>
```

Note that this code represents only the part of the CCD ontology that correspond the Medications section of the CCD. The whole CCD ontology is comprised of the integration of the ontologies derived from all sections of the CDD. In such integration we do not have to take care of semantic heterogeneity (i.e., one term is used in different meanings, or two terms are used in same meaning) as the all the elements in CCD documents are based on the RIM.

In order to illustrate the integration, the graphical ontology in Fig. 8 represents a portion of the CCD ontology. It includes elements from the Medications section and from the Vital signs section. Note that this ontology contains sufficient information for specifying the query presented in Section 1, i.e., "give me the average blood pressure and/or cholesterol level during the time periods a patient, say Tim Jones, was using Diovan".

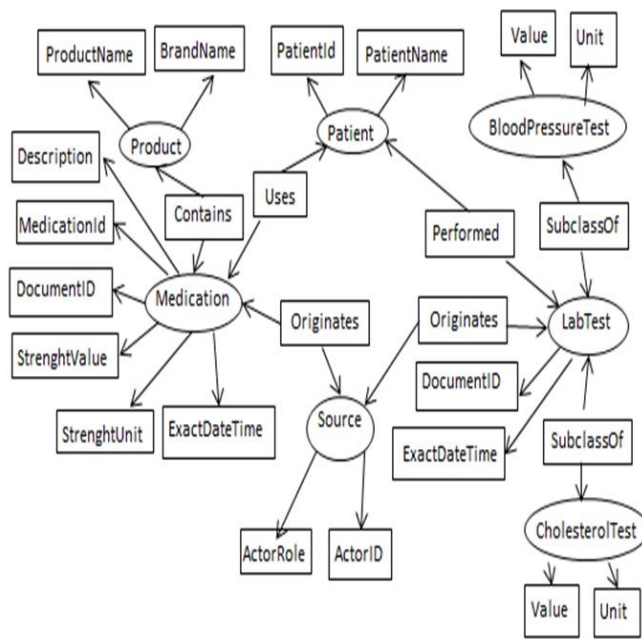


Figure 8. A portion of the CCD ontology

B. Querying the CCD ontology by RQL and SPARQL

The content of a clinical document is more than just the sum of the individual facts and suppositions stored inside the document meaning that each statement in the document relates to other statements contained in the document [11]. For example, the statement about the medication is important but may not be fully understood without looking at the particular diagnosis, allergies or recorded intolerances. So the information contained within a document is expected to be understood in the context of the whole.

Clinical documents also have specific properties that are not shared with traditional databases [12]. An essential requirement is that the original documents can be generated from the data stored in the database or data archive [13]. For example, the data items stored in the CCD ontology of Fig.4 are originated either from a medication document or from a laboratory test document (or these both may be included in the same CCD document, meaning that they both have the same value on the attribute DocumentID), and thereby it is necessary that these documents can be reconstructed later on.

Reconstructing original documents (or representing queries) by RQL and SPARQL on the EHR-Archive is rather easy. For example, in RQL to retrieve all instances of the class Medication (i.e., all Medication documents) we only have to write "Medication".

To retrieve all medications of the patient having value AB-12345 on PatientID (i.e., Tim Jones) we have to write the following query:

```
select N
from Patient{X}.uses{Y},{C} MedicationID {N}
where X="AB-12345" and X=C
```

of course the physicians do not have to be familiarized with query languages in order to retrieve data from the EHR-Archive as user-friendly interfaces can be easily developed.

VIII. INTEROPERABILITY IN ARCHIVING CLINICAL DOCUMENTS

In our used architecture the original patients' EHRs are stored in healthcare providers' EHR systems, and they are the data sources for the EHR-Archive [14] (Fig. 13). EHR systems are a part of a local stand-alone health information system that allows storage, retrieval and modification of records [15]. EHR-Archive is managed by healthcare authorities.

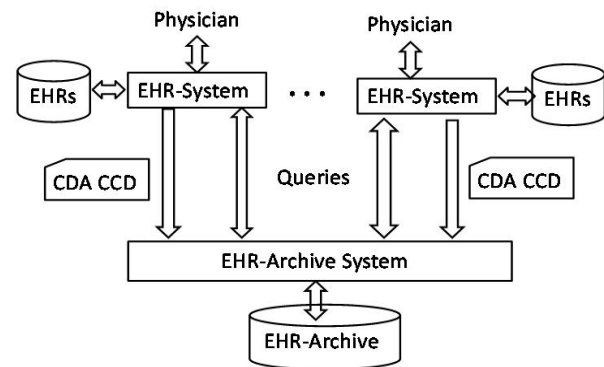


Figure 9. EHR-Archive System.

As illustrated in Fig. 9 the physicians query the EHR-Archive through their used EHR systems, and the medical information transmitted to the EHR-Archive are CDA CCD documents. The EHR-Archive system transforms CDA CCD documents into format that is compliant with the CCD ontology. Such a transformation is illustrated in Fig. 10.

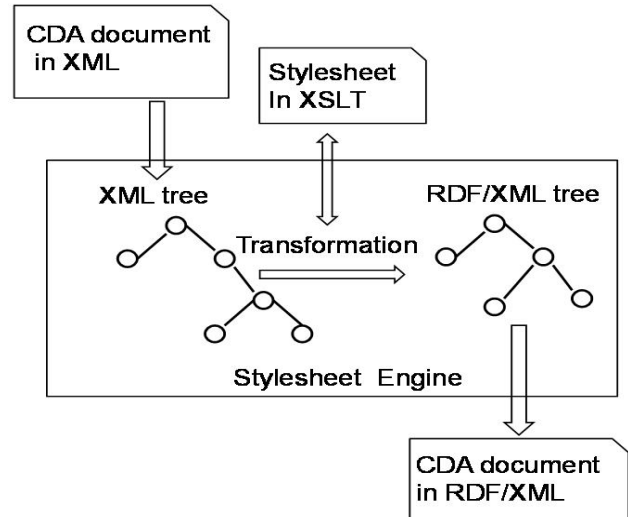


Figure 10. Transforming a CDA document into RDF/XML format.

The Stylesheet Engine takes an XML document, loads it into a DOM (Document Object Model) source tree, and transforms that document with the instructions given in the style sheet into RDF/XML format. The instructions use XPath [12] expressions in referencing to the source tree and in placing it into the result tree. The result tree is then formatted, and the resulting element in RDF/XML format is returned.

After the XML document is transformed to appropriate RDF/XML element it is inserted to the CCD ontology. The CDA CCD document presented after Fig. 3 is presented in transformed RDF/XML format as follows:

```

<rdf:RDF
  xmlns : rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns : xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns : po="http://www.lut.fi/ontologies/CCD-Ontology#"
  <rdf:Description rdf:about="AB-12345">
    <rdf:type rdf:resource="&po;Patient"/>
    <po : PatientName>Tim Jones</po : PatientName>
    <po : Uses>MO-5481</po:Uses>
    <po : Performed>H-257L</po : Performed>
  </rdf : Description>
  <rdf:Description rdf:about=" MO-5481">
    <rdf:type rdf:resource="&po;Medication"/>
    <po : Contains>Voltaren</po : Contains>
    <po:ExactDateTime>2012-03-01T012:00
    </po : ExactDateTime>
    <po : StrenghtValue rdf:datatype=
      "&xsd:integer">30</po : StrenghtValue>
    <po : StrenghtUnit>Tabs</po : StrenghtUnit>
  </rdf : Description>
  <rdf:Description rdf:about=" 211708-8">
    <rdf:type rdf:resource="&po;Source"/>
    <po : ActorID>Pharmacy</po : ActorID>
    <po : ActorRole>Pharmacy</po : ActorRole>
  </rdf : Description>
  <rdf:Description rdf:about=" Voltaren">
    <rdf:type rdf:resource="&po;ProductName"/>
    <po : BrandName>Diclofenac</po : Contains>
  </rdf : Description>
</rdf:RDF>

```

The above RDF/XML-formatted document of is comprised of four RDF-descriptions. Further, the first RDF-description is comprised of four RDF-statements. The first statement states that the type of the instance identified by "AB-12345" is Patient in the EHR-Ontology. The second RDF-statement states that the name of the instance identified by "AB-12345" is Tim Jones.

IX. CONCLUSIONS

Interoperability in healthcare means the ability of two or more healthcare systems to exchange clinical documents and to use the information of the exchanged documents. The Clinical Document Architecture (CDA) is an ANSI approved HL7 standard. It is proven to be a valuable and powerful standard for a structured exchange of clinical documents between healthcare information systems. However, though healthcare systems interoperability solves the problem of patients' scattered health documentation, it does not solve the problem of querying the content of patients' health documentation.

In order to solve this problem we have studied the suitability of the Semantic Web technologies for integrating patients' health documentation. In particular, we have developed two ontology-based methods for integrating the content of CCD documents. Which of the methods is appropriate depends on whether the header or the whole documents are based on the HL7 RIM.

In our future work we will study the suitability of cloud computing for the EHR-Archive system. Cloud computing represents new way of delivering organization information technology: anyone with a suitable Internet connection and a standard browser can access an application in a cloud. However, in spite of the widespread adoption of cloud

computing by most industries, the healthcare sector has been rather slow in adopting cloud-based solutions. This slow adoption is partially due to concerns about data security and compliance with key regulations.

Assuming that data security and compliance with key regulations are met, we assume that cloud computing will provide significant benefits to healthcare organizations and help them improve patient care. It also allows new healthcare delivery models that will make healthcare more efficient and effective. However, the success of new cloud-based technologies mainly depends on how they can be adopted to prevailing healthcare infrastructures.

In addition, there are some risks that may jeopardize the success of new technology. Especially the introduction new technology requires training: the incorrect usage of a new e-health technology, due to lack of proper training, may ruin the whole system. Also a consequence of introducing a new healthcare practice is that it significantly changes the daily duties of the healthcare personnel. Therefore one challenging aspect is also the changing the mind-set of the involved healthcare personnel.

REFERENCES

- [1] Benson, T., Principles of Health Interoperability HL7 and SNOMED. Springer. 2010.
- [2] Agarwal, R, Angst, C.M., Technology-enabled transformations in U.S. health care: early findings on personal health records and individual use," In Galletta G, Zhang P, (eds.), Human-Computer Interaction and Management Information Systems: Applications (Vol. 5). Armonk, NY: M.E. Sharpe, Inc., pp. 357-378. 2006.
- [3] Lewis, D., Eysenbach, G., Kukafka, R., Stavri, P.Z, Jimison, H.: Consumer health informatics: informing consumers and improving health care. Springer. 2005.
- [4] Angst, C.M., Agarwal, R, Downing, J., An empirical examination of the importance of defining the PHR for research and for practice, In: 41st Annual Hawaii International Conference on System Sciences, 2008.
- [5] Kaelber, D., Jha, A., Johnston, B., Middleton, B., Bates, D.: A Research Agenda for Personal Health Records (PHRs). J. Am. Med. Inform. Assoc. 15(6), 729 – 736. 2008.
- [6] Raisinghani M.S., Young, E., Personal health records: key adoption issues and implications for management, International Journal of Electronic Healthcare. Vol. 4, No.1 pp.67-77. 2008.
- [7] Puustjärvi J., Puustjärvi L., Managing Medicinal Instructions. In the proc. of the International Conference on Health Informatics (HEALTHINF 2009). pp. 105-110. 2009.
- [8] Liu. S., Personal Health Records: Status-Quo and Future Perspectives, in Ubiquitous Health and Medical Informatics". Ed. S. Mohammed and J. Fiadhi, IGI-Global, pages 43-63. 2010.
- [9] Puustjärvi J., Puustjärvi L., The role of medicinal ontologies in querying and exchanging pharmaceutical information. International Journal of Electronic Healthcare, Vol. 5, No.1 pp. 1 – 13. 2009.
- [10] Puustjärvi, J., Puustjärvi, L., Personal Health Ontology: Towards the Interoperation of E-health Tools. International Journal of Electronic Healthcare. Vol. 6, No.1, 62 – 75. 2011.
- [11] Boone, K., The CDA Book. Springer.2011.
- [12] Puustjärvi, J., Puustjärvi, L., Personal Health Book: A Novel Tool for Patient Centered Healthcare. In the Proc of the

- International Conference on Health Informatics (HEALTHINF 2011), pp. 386-393. 2011.
- [13] Puustjärvi, J., Puustjärvi, L., Moving from Remote Patient Monitors to Cloud-Based Personal Health Information Systems: a Way to Practicing Patient-Centered Chronic Care Model. In the proc. of the International Conference on Health Informatics (HEALTHINF 2012)). Pages 37-45. 2012.
- [14] Puustjärvi, J., Puustjärvi, L., Towards Semantic Exchange of Clinical Documents, International Journal on Advances in Life Sciences (IJALS). Vol. 1, No.2&3 pp. 69 – 76 .2009.
- [15] Puustjärvi, J., Puustjärvi, L., Ontology-Based Electronic Health Record Archive: A Way to Content-Based Querying of Patient's Clinical Documentation. Second International Conference on Advances of Information & Communication Technology in Health Care – ICTHC 2012. 2012.